CHAPTER

4

Molecular Modeling of Major Structural Protein Genes of Avian Coronavirus: Infectious Bronchitis Virus Mass H120 and Italy02 Strains

Khadija Khataby[1,2], *Yassine Kasmi*[1], *Amal Souiri*[1], *Chafiqa Loutfi*[2] *and Moulay Mustapha Ennaji*[1]

[1]**Laboratory of Virology, Microbiology, Quality, Biotechnologies/ Eco-Toxicology and Biodiversity, Faculty of Sciences and Techniques, Mohammedia, University Hassan II of Casablanca, Casablanca, Morocco**

[2]**Society Biopharma, Rabat, Morocco**

INTRODUCTION

Avian infectious bronchitis (IB) is a highly contagious respiratory infectious disease hazardous to the poultry industry. It can infect chickens at all ages and replicates in many tissues, causing respiratory symptoms, diarrhea, decline of egg production and quality, etc. (Cavanagh, 2007a,b; Abd et al., 2009). Prevention of IB is of economic importance to the poultry industry due to the high morbidity and production losses associated with the disease (Cavanagh, 2005). Although vaccines are now being used widely and extensively, outbreaks of IB still occur frequently due to the epidemic IB virus (IBV) strains (Zou et al., 2010). It is well known that little or no cross-protection occurs between different

Emerging and Reemerging Viral Pathogens
DOI: https://doi.org/10.1016/B978-0-12-814966-9.00004-4

serotypes of IBV, and new serotypes may appear in the future, complicating the prevention and control of IB.

In Morocco, the epidemiological situation of IBV is very complex due to the antigenic diversity associated with the emergence of new serotypes/genotypes and variants, vaccination failures linked to a possible maladjustment of the vaccine strain used and/or poor vaccination practices, and inadequate biosecurity measures by livestock keepers. The avian IBV strains in circulation are serotypes/genotypes Italy02 and Mass H120 identified since 2010 (Fellahi et al., 2015).

The etiologic agent of IB is IBV, a prototype of the *Coronaviridae* family, which is an enveloped, positive sense, single-stranded RNA virus (Boursnell et al., 1987).

The viral genome is around 27.6 kb in length and encodes four structural proteins, nucleocapsid protein (N), membrane glycoprotein (M), spike glycoprotein (S), and small envelope protein (E) (Lai et al., 1981). The S glycoprotein is posttranslationally cleaved at protease cleavage recognition motifs into the animal-terminal S1 and carboxyl-terminal S2 subunits by cellular protease (Jackwood et al., 2001; Cavanagh et al., 1986). The S1 glycoprotein contains epitopes that induce virus-neutralizing, serotype-specific antibodies, hemagglutination inhibition antibodies, and cross-reactivity enzyme-linked immunosorbent assay (ELISA) antibodies (Niesters et al., 1987). It also plays an important role in tissue tropism and the degree of virulence of the virus (Casais et al., 2003). The appearance of these variants hinders the prophylactic strategy carried out by the breeders of the Moroccan poultry farm. In order to solve this problem, we have opted to study the structure of the hypervariable region of the S1 protein of serotype Italy02 and Mass in silico by molecular modeling, where the largest number of epitopes identified by neutralizing antibodies is observed (Koch et al., 1992).

Structural bioinformatics is a branch of bioinformatics that focuses on the prediction of macromolecular structures, such as the structure of three-dimensional (3D) proteins (Zhang et al., 2005).One of the main questions in the problem of protein structure prediction is the challenge of understanding how the primary protein structure information is translated into a 3D structure and how to use this information for the development of prediction of the 3D structure (Creighton, 1990).

Experimental methods for determining the 3D structure of proteins are cumbersome and costly in terms of time and resources. The predictive methods in silico propose a fast and efficient alternative, based on a set of physical, statistical, and biological laws. Generally, there are two main classes of methods: the first class is called “comparative modeling” and the second is called “ab initio” (Piuzzi, 2010). The first method depends on the existence of homologous proteins, whose structures are

determined experimentally. The second method is only on physical and statistical laws. The algorithms used by the latter are very greedy in computing time, and the results obtained progress with advances in computer science.

Actually, despite the immense progress of ab initio methods, comparative methods are still those which offer the best predictions of antigenic sites of proteins. Therefore the objective of the present study, which is reported for the first time in Morocco, aims to compare the structural conformation of the S1 protein in 3D form, and to predict the common neutralizing epitopes between the vaccine strain Mass H120, which is the most dominant serotype in Morocco, and the serotype Italy02 to better understand their pathogenic and immunogenicity.

MATERIAL AND METHODS

Sequence and Structural Data

The viral strains used in this study were Italy02 and Mass (H120). The amino acid sequences for the proteins to be modeled were obtained from Genbank NCBI-USA, and their access numbers are (KM594188: Italy02) and (M21970: H120). The evolutionary characterization of IBV is essentially based on the analysis of the three hypervariable regions of the S1 gene (HVR1, HVR2, and HVR3), located in the following positions: 114−201nt, 297−423nt, and 822−1161nt, respectively, corresponding to amino acid residues 38−67, 91−141, and 274−38 (Bourogâa et al., 2009).

MODELING OF THE HYPERVARIABLE REGION OF S1 SPICULE PROTEINS

This paper focuses on the molecular modeling of the structure of the hypervariable regions of the S1 protein of the Mass H120 and Italy02 serotypes circulating in Morocco. To meet the stated objective, an alignment of the protein sequences of these two strains was carried out in order to detect the homologous and common regions, so that it could be applied to the model described by Piuzzi (2010). All these manipulations were developed by The CHIMERA V.01 software.

Homology modeling allows replacing the missing structural information, provided that it has the structure of a protein with strong homology between the two sequences “Italy02 and H120.” It is estimated that two structures can be considered identical when their root,

mean-square, deviation (RMSD) (obtained by the superposition of the atoms of their respective main chains) is less than 2 Å.

In this study, modeling was done using the I-TASSER server, then the COACH, and another Meta server to determine and predict the common immunogenic active sites, between these two IBV strains. The 3D modeling was carried first, with I-TASSER, which is a server offering a service of prediction of the structure and function of the protein studied. It makes it possible to produce high-quality 3D models from the amino acid sequences. The results provided by the I-TASSER server are in the form of several 3D models, classified according to a score called "TM-score." If the TM-score is greater than 0.5, it indicates that the model generated a valid topology. However, a score less than 0.17 indicates a random topology (Piuzzi, 2010).

After the 3D modeling of the hypervariable regions, the next step was to calculate and determine the active site of the modeled regions, the site where the ligand interaction takes place, which results in activation or deactivation of the biological function of the protein.

The calculation of the active site was carried out with COACH which is a meta server for the prediction of the "ligand binding domain." The 3D models generated by I-TASSER are taken into account by the COACH server to predict all active sites with their ligands. The active sites obtained by COACH are coordinates in this form of "X/Y/Z/Xs/Ys/Zs" or "X, Y, and Z" represent the position of the active site in 3D space and for "Ys and Zs" show the size of the box containing the immunogenic active site.

Another server was used, named "RAMACHANDRAN," giving information on the conformation of the protein in 3D. Thanks to the diagrams generated by this server, the potential secondary structures can be identified according to the torsion angles. There are two types of torsion angles, the angle phi (φ) and the angle psi (ψ). The angle psi (Ψ) represents the angle of rotation around the Cα−C bond (of C = O) of the plane 1 and the angle phi (φ) represents the angle of rotation around the Cα−N bond (of NH) of the plane 2.

EVALUATION AND REFINEMENT OF THE THREE-DIMENSIONAL MODEL

The quality and precision of the models obtained were evaluated by the geometry of the different regions of the model and the identification of possible errors. The evaluation of the quality of the 3D modeled protein was carried out by the "PROSESS: Protein Structure Evaluation Suite & Server." PROSESS is a web server designed to

evaluate and validate protein structures and allows us to integrate a variety of analyzes:

- covalent and geometric quality
- noncovalent bond quality
- quality of the torsion angle
- chemical shift quality

PROSESS produces detailed tables with explanations, images, and graphs that summarize the results by comparing them with values observed in high-quality protein structures. This server is used to coordinate the location of hydrogen bonds, secondary structure, and geometric analysis, which can then be used for computation of aliasing and solvent energy, and chemical shift correlations, to correlate the mobility of the structure with chemical shift, as well as for the calculation of torsional angle and chemical changes (Berjanskii et al., 2010).

RESULTS

The study of the spatial conformation of the structure of the hypervariable region of the S1 protein in 3D and the prediction of the neutralizing epitopes of the virus were carried out using tools of molecular modeling.

MODELING OF THE HYPERVARIABLE REGION OF S1 SPICULE PROTEINS

Spatial Conformation of the S1 Structure in Three-Dimensional

Homology modeling between the two protein sequences (Italy02 and Mass) showed a similarity percentage of 81%. This homology was evaluated by the I-TASSER server, which allows us to generate 3D models from the protein sequence. These models were then ranked in specific order and defined by the TM-score which measures the deviation distance (Angstrom) between the residual position of the model and the native structure. The score obtained by I-TASSER revealed that the two modeled sequences used in this study have a more significant TM-score that exceeds the value of 0.5 (TM-Score = 0.63), confirming that the model is biologically significant and has a correct structural topology (Fig. 4.1, Table 4.1).

Projection and disposition of the 3D structure of the Italy02 strain on that of the Mass strain was validated by the RMSD factor. This factor evaluates the degree of deviation between the two 3D structures.

The RMSD is equal to 0.3 Å, whereas most hypervariable regions have an RMSD equal to zero, indicating that both structures are

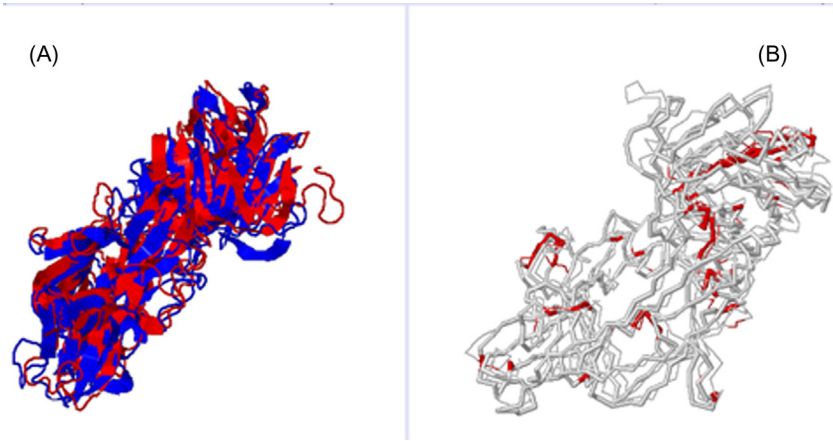


FIGURE 4.1 (A) 3D structure predicted by I-TASSER; Blue color: strain H120. Red color: strain Italy02. (B) The common regions between the two structures (red). *3D*, Three-dimensional. (*Note: For interpretation of the references to color in this figure legend, the reader is referred to the web version of this chapter*).

TABLE 4.1 Scores TM and C As Well As Predicted Active Sites and Exogenous Molecules

Sequences	C score	TM-score	RMSD	Active sites	Ligand
Italy02	−0.91	0.60 ± 0.14	9.6 ± 4.6 Å	169, 171, 179, 208, 224, 229, 232, 233, 234, 235, 236, 237	Mg^{2+}
H120	−0.91	0.60 ± 0.14	9.6 ± 4.6 Å	225, 228, 229, 230, 232, 233, 235, 338, 341, 433, 435, 440, 476, 485	Mg^{2+}

RMSD, Root, mean-square, deviation.

identical and share homologous and common regions (Fig. 4.1B). In addition, both strains share common active sites in the S1 spike protein and are located at residues 229, 230, 232, 233, and 235 (Table 4.1). This study also revealed the presence of a molecule of magnesium associated with the structure of the amino acids common between the two sequences of the strains studied.

STABILITY OF THE STRUCTURE OF THE S1 PROTEIN IN THREE-DIMENSIONAL

In order to confirm the stability of the 3D structure, several Beta and Alpha sheets were demonstrated by RAMACHANDRAN test. The analysis of these results showed a variability in the stability of the

sequences, depending on the number of residues outside the stability zones. The amino acids are distributed between Beta sheets and Alpha helices (Fig. 4.2, Table 4.2).

The amino acids distributed in the upper left quadrant indicate those found in the Beta leaflets. They have an angle Phi less than −30 and an angle psi greater than 90. The set of proteins in the white space is of suspended structure and of unknown nature (Fig. 4.2).

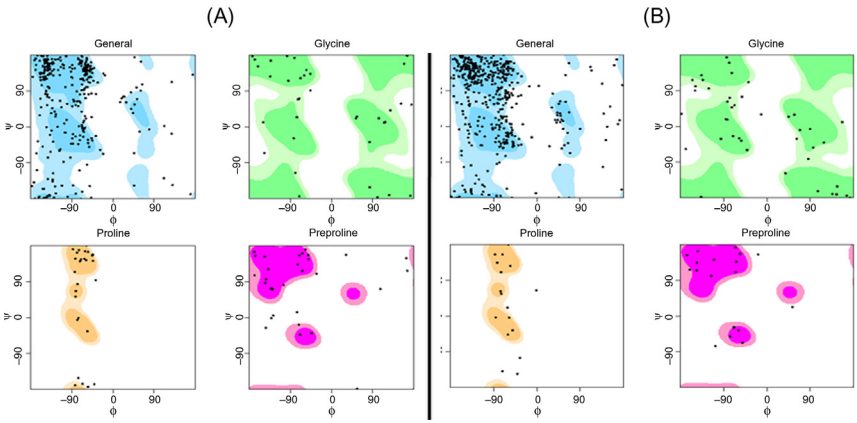


FIGURE 4.2 RAMACHANDRAN plot (A) H120 and (B) Italy02.

TABLE 4.2 Results of RAMACHANDRAN Analysis

Residues types	Percentage of outliers	Number of percentage of outliers	Total number
ITALY02			
Générale (non-Gly, non-Pro, non-pre-Pro)	12.58	39	310
Glycine	10.26	4	39
Proline	19.35	6	31
Preproline	22.58	7	31
H120			
General (non-Gly, non-Pro, non-pre-Pro)	10.43	48	460
Glycine	13.33	6	45
Proline	31.58	6	19
Preproline	15.79	3	19

The arrangement of the amino acids in the lower left part coinciding on the one hand, with the right-handed alpha-helix conformation, and on the other hand, a small number of amino acids is located in the upper right quadrant, By showing alpha helices rotating to the left through their conformation angles. They also have an average stability of between 10 and 12 (Fig. 4.2).

EVALUATION OF THE QUALITY OF THE THREE-DIMENSIONAL MODEL AND PREDICTION OF ANTIGENIC SITES

Evaluation of the Quality of the Three-Dimensional Model

The analysis of the evaluation of the structural quality of 3D sequences by the PROSESS server showed that they have an overall quality of 2.5. All residues, within the range of $20 < R < 60$ are characterized only by noncovalent bonds with a value of two anomalies, while residues included in $74 < R < 500$ are indicated by the packing and noncovalent bonds, whose average quality is equal to 3.5, then for the other covalent bonds, it reaches a value of 4.5 (Fig. 4.3).

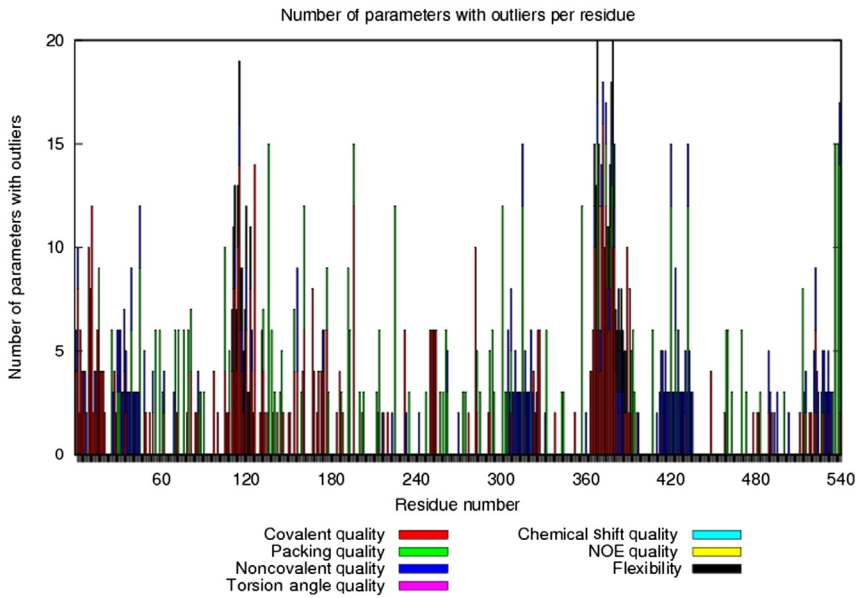
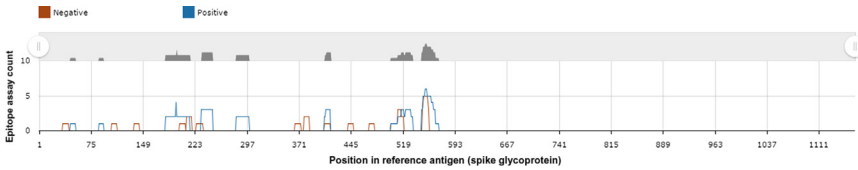


FIGURE 4.3 **Aggregated results of sequence residue problems.**

PREDICTION OF ANTIGENIC SITES

The results of the prediction of neutralizing epitopes at the level of the S1 protein in 3D showed that the serotype Italy02 and the vaccine strain H120 of serotype Mass share common epitopes in the hypervariable regions of the S1 spicule protein, which may have antigenic and immunogenic role. These epitopes are at residues 38–67, 91–141, and 274–387



Prediction of epitopes at the spicule protein structure S1.

DISCUSSION

This study would focus on the in silico prediction of peptides of the S1 spicule protein from two IBV strains, Italy02 and Mass H120. The choice of the S1 subunit was not made by chance but was chosen by its ability to undergo mutations in the hypervariable regions, giving new strains of IBV (Cavanagh, 2007a,b). The S1 subunit anchors to the outer surface of the viral particle, making it the more easily recognized antigen, by the IB-specific antibody, compared to other IBV antigens.

The S1 gene is now commonly used as a marker of the IBV classification. Although it is highly variable, it remains the first choice for the development of subunits of vaccines against IB (Zou et al., 2015). Furthermore, since there are still relatively conserved regions or epitopes in the S1 subunit, S1 could also be used as a targeted antigen in the development of diagnostic agents (Zou et al., 2015). However, there is little information on the structure of the S1 gene protein in 3D, which is carried out to predict epitopes on this gene, hence the objective of this paper, which aimed to predict and identify the most immunogenic antigenic sites are critical for vaccine development.

The results of the homology modeling showed that the two studied serotypes had almost the same spatial conformation of the hypervariable region of the S1 protein in 3D and shared homologous and common regions with a similarity percentage of 81%. These results are in agreement with the data reported by Chothia and Lesk (1986). These authors

have shown that for RMSD values below 2 Å, the two structures can be considered as similar. Thus from 60% homology, homology modeling allows correct prediction in 70% of cases (Chothia and Lesk, 1986).

Analysis of the results of structural stability showed that there were residual stability and variability between the two protein sequences (Italy02 and Mass) in 3D. Jones and Jordan (1972) reported that the serotype Mass H120 is more stable than the serotype Italy02, whose ratio of the predicted strains is 1.5. These data could be explained by the evolutionary power of this virus as a function of time, where mutation occurs at a speed faster than normal in the hypervariable sequence of the S1 gene, which is the subject of this study.

The study of the prediction of epitopes revealed the presence of common active residues at the level of the hypervariable region of the spike protein S1, which can exercise a common function by intervening in the juice of internalization of the virus of the cell, thus the role of cathepsins.

In addition, detection of a magnesium molecule was detected associating with the structure of amino acids around Aln280, a common predicted region and considered to be one of the most immunogenic regions in both IBV strains.

The presence of the magnesium molecule around this site stimulates immunogenicity, which has been researched because of its functionality in the body (Tam et al., 2003). These authors have shown that this combination site antigen−magnesium has a strong relationship with the immune system, both in the nonspecific and specific immune response, also called innate and acquired immune response (Tam et al., 2003). That is, as a cofactor for the synthesis of immunoglobulins, C′3 convertase, immune cell adhesion, antibody-dependent cytolysis, IgM lymphocyte binding, macrophage response to lymphokines, and the adhesion of helper T lymphocytes (Tam et al., 2003). These data are in accordance with the results described above (Zou et al., 2015). These authors have demonstrated that this molecule promotes more antigenic and immunogenic power around this active site, giving an immune response of 100% neutralizing antibodies. The reason might be that despite significant differences in the S1 protein, much of the virus genome remains unchanged, and there are common epitopes among different strains of IBV, which play a major role in protective immunity (Cavanagh, 1997).

Based on the results presented here, the two protein sequences studied have a 3D spatial conformation and common predicted neutralizing epitopes, where it seems that both strains have the same pathogenicity and tissue tropism.

Thus it is highly probable that the H120 vaccine strain confers cross-protection against a challenge with new strains Italy02 circulating in Morocco. So far, Mass strains have been mainly used as live vaccines because of their epizootic distributions and cross-protective capacity (Ignjatovic and Sapats, 2000, Bijlenga et al., 2004).

CONCLUSION

The in silico study presented here shows that the two serotypes Italy02 and Mass H120 circulating in Morocco share an identical structure in 3D, with a similarity percentage of 81%, as well as common predicted neutralizing epitopes. To realize this data, experimental research on the cross-protection between the two serotypes detected in this country is necessary.

COMPETING INTERESTS

Authors declare they have no competing interests and had access to all generated data and that they contributed to the analyses and interpretation.

References

Abd, E.R.S., El-Kenawy, A.A., Neumann, U., Herrler, G., Winter, C., 2009. Comparative analysis of the sialic acid binding activity and the tropism for the respiratory epithelium of four different strains of avian infectious bronchitis virus. Avian Pathol. 38 (1), 41–45.

Berjanskii, M., Liang, Y., Zhou, J., Tang, P., Stothard, P., Zhou, Y., et al., 2010. PROSESS: a protein structure evaluation suite and server. Nucleic Acids Res. 38, W633–W640.

Bijlenga, G., Cook, J.K.A., Gelb, J., de Wit, J.J., 2004. Development and use of the H strain of avian infectious bronchitis virus from the Netherlands as a vaccine: a review. Avian Pathol 33, 550–557.

Boursnell, M.E., Brown, T.D., Foulds, I.J., Green, P.F., Tomley, F.M., Binns, M.M., 1987. Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus. J. Gen. Virol. 68, 57–77.

Casais, R., Dove, B., Cavanagh, D., Britton, P., 2003. Recombinant avian infectious bronchitis virus expressing a heterologous spike gene demonstrates that the spike protein is a determinant of cell tropism. J. Virol. 77 (16), 9084–9089.

Cavanagh, D., 1997. *Nidovirales:* a new order comprising *Coronaviridae* and *Arterividae*. Arch. Virol 142 (3), 629–633.

Cavanagh, D., 2005. Coronaviruses in poultry and other birds. Avian Pathol. 34 (6), 439–448.

Cavanagh, D., 2007a. Coronavirus avian infectious bronchitis virus. Vet. Res 38, 281–297.

Cavanagh, D., 2007b. Coronavirus avian infectious bronchitis virus. Vet. Res. 38 (2), 281–297.

Cavanagh, D., Davis, P.J., Darbyshire, J.H., Peters, R.W., 1986. Coronavirus IBV: virus retaining spike glycopolypeptide S2 but not S1 is unable to induce virus-neutralizing or haemagglutination-inhibiting antibody, or induce chicken tracheal protection. J. Gen. Virol. 67 (Pt 7), 1435–1442.

Chothia, C., Lesk, A.M., 1986. The relation between the divergence of sequence and structure in proteins. EMBO J 4, 823–826.

Fellahi, S., Ducatez, M., El Harrak, M., Guerin, J.L., Touil, N., Sebbar, G., et al., 2015. Prevalence and molecular characterization of avian infection bronchitis virus in poultry

flocks in Morocco from 2010–2014 and the first report of Italy02 genotype in Africa. Avian Pathol. 44, 287–295.

Ignjatovic, J., Sapats, S., 2000. Avianinfectiousbronchitis virus. Rev. Sci. Tech 19, 493–508.

Jackwood, M.W., Hilt, D.A., Callison, S.A., Lee, C.W., Plaza, H., Wade, E., 2001. Spike glycoprotein cleavage recognition site analysis of infectious bronchitis virus. Avian Dis. 45 (2), 366–372.

Jones, R.C., Jordan, F.T., 1972. Persistence of virus in the tissues and development of the oviduct in the fowl following infection at day old with infectious bronchitis vims. Res. Vet. Sci. 13, 52–60.

Koch, A.E., Polverini, P.J., Kunkel, S.L., Harlow, L.A., DiPietro, L.A., Elner, V.M., Strieter, R.M., 1992. Interleukin-8 as a macrophage-derived mediator of angiogenesis. Science 258 (5089), 1798–1801.

Lai, M.M., Brayton, P.R., Armen, R.C., Patton, C.D., Pugh, C., Stohlman, S.A., 1981. Mouse hepatitis virus A59: mRNA structure and genetic localization of the sequence divergence from hepatotropic strain MHV-3. J. Virol. 39 (3), 823–834.

Niesters, H.G., Kusters, J.G., Lenstra, J.A., Spaan, W.J., Horzined, M.C., van der Zeijst, B. A., 1987. The neutralization epitopes on the spike protein of infectious bronchitis virus and their antigenic variation. Adv. Exp. Med. Biol. 218, 483–492.

Piuzzi, M., 2010. Détermination de la structure de protéines à l'aide de données faiblement résolues (Thèse Biochimie). Université Pierre et Marie Curie, Paris VI, Français.

Tam, M., Gómez, S., González-Gross, M., Marcos, A., 2003. Possible roles of magnesium on the immune system. Eur. J. Clin. Nutr 57, 1193–1197.

Zhang, Y., et al., 2005. Cloning, expression and characterization of the human NOB1 gene. Mol Biol Rep. 32 (3), 185–189.

Zou, N.L., Zhao, F.F., Wang, Y.P., Liu, P., Cao, S.J., Wen, X.T., et al., 2010. Genetic analysis revealed LX4 genotype strains of avian infectious bronchitis virus became predominant in recent years in Sichuan area, China. Virus Genes 41 (2), 202–209.

Zou, N.L., Wang, F.F., Duan, Z., Xia, J., Wen, X., Yan, Q., et al., 2015. Development and characterization of neutralizing monoclonal antibodies against the S1 subunit protein of QX-like avian infectious bronchitis virus strain Sczy3. Monoclon. Antib. Immunodiagn. Immunother. 34, 17–24.